ABSTRACT

An utterance detector for speech recognition is described. The detector consists of two components. The first part makes a speech/non-speech decision for each incoming speech frame. The decision is based on a frequency-selective autocorrelation function obtained by speech power spectrum estimation, frequency filter, and inverse Fourier transform. The second component makes utterance detection decision, using a state machine that describes the detection process in terms of the speech/non-speech decision made by the first component.